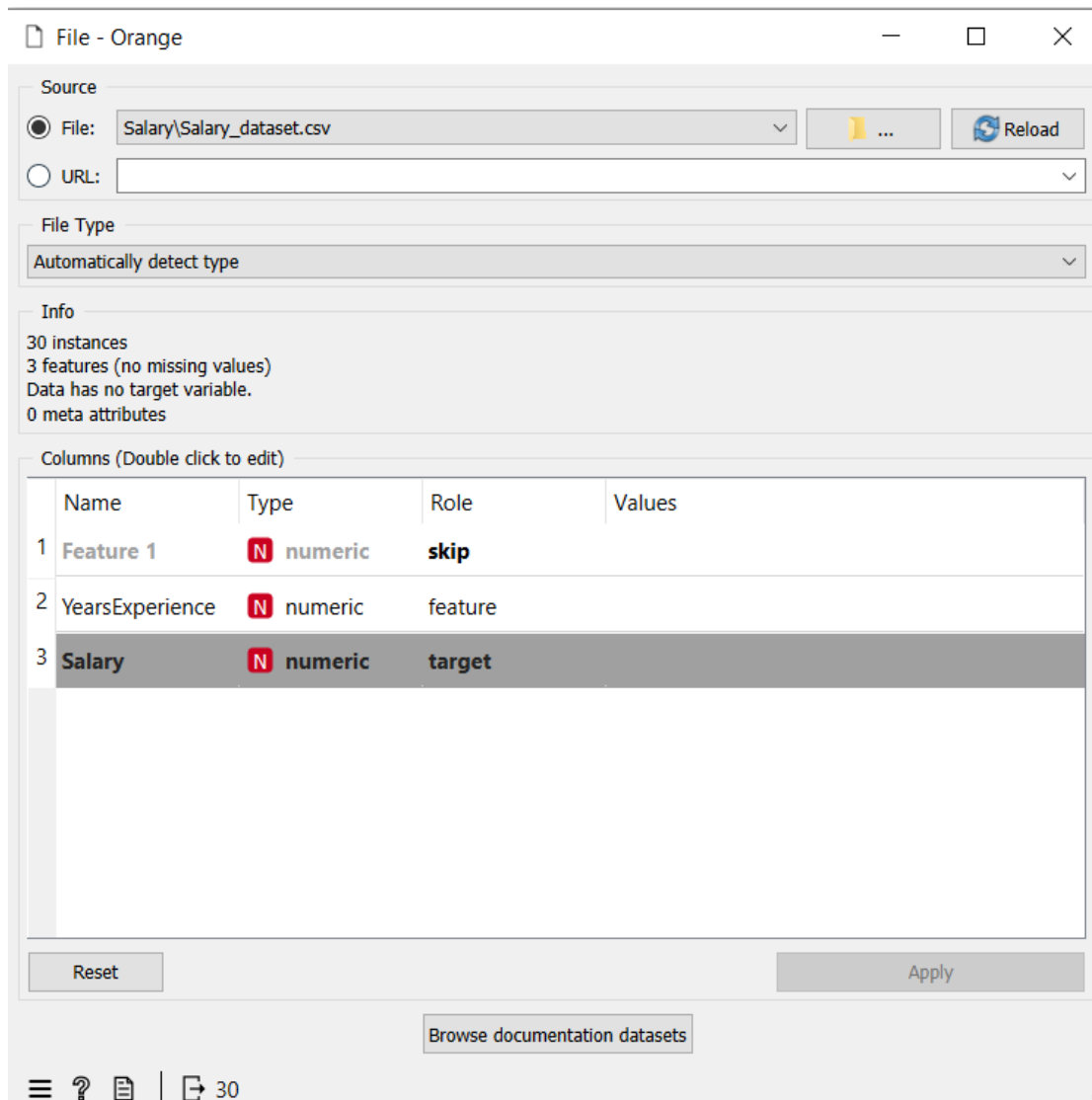


Diagram tersebut menggambarkan alur proses analisis data yang dimulai dari pengunggahan file. Data yang diimpor kemudian disusun dalam bentuk tabel, yang memungkinkan pengguna untuk menganalisis statistik fitur dan membuat visualisasi seperti box plot dan scatter plot untuk memahami distribusi dan hubungan antar variabel. Selanjutnya, data dapat diambil sampelnya menggunakan Data Sampler untuk pengujian lebih lanjut. Proses ini mengarah pada penerapan model regresi linier, di mana data yang telah disiapkan digunakan untuk melatih model. Setelah pelatihan, model dievaluasi melalui proses pengujian dan penilaian untuk menghasilkan prediksi. Diagram ini memberikan gambaran menyeluruh mengenai langkah-langkah dalam analisis data, dari pengolahan awal hingga pembuatan prediksi akhir menggunakan model statistik.



Gambar tersebut menunjukkan antarmuka impor data di Orange, sebuah alat visualisasi dan analisis data. File yang diimpor bernama "SalarySalary\_dataset.csv" dan berisi 30 instansi, tanpa nilai yang hilang. Dataset ini terstruktur dengan dua kolom: "Feature 1," yang diberi label sebagai fitur numerik (dengan peran "skip"), dan "Salary," yang ditetapkan sebagai variabel target dan juga numerik. Pengaturan ini menunjukkan bahwa dataset kemungkinan ditujukan untuk tugas pemodelan prediktif, di mana tujuannya adalah memprediksi gaji berdasarkan fitur yang disediakan. Pengguna dapat menerapkan perubahan atau mengatur ulang pengaturan sebelum melanjutkan dengan analisis lebih lanjut. Terdapat dua kolom yang dijelaskan sebagai berikut:

Feature 1:

Tipe: Numerik

Peran: Skip

Keterangan: Kolom ini berisi data numerik yang tidak akan digunakan dalam analisis atau pemodelan lebih lanjut karena perannya adalah "skip." Ini berarti kolom ini mungkin tidak relevan untuk prediksi atau analisis yang diinginkan.

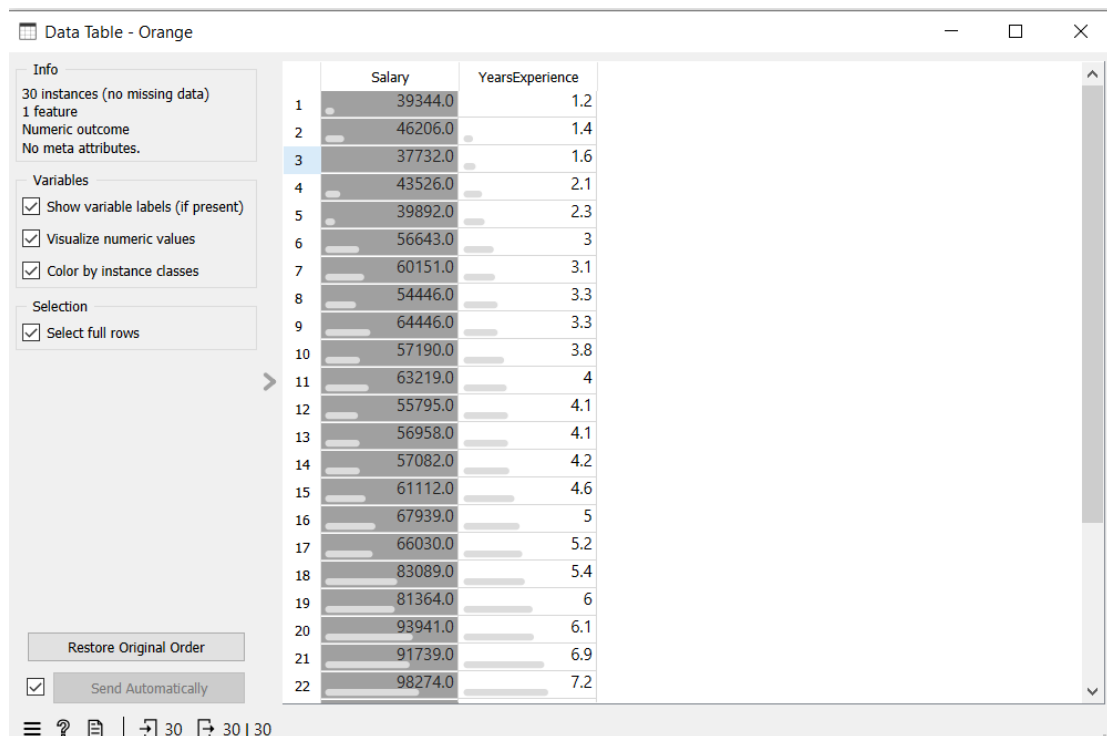
Salary:

Tipe: Numerik

Peran: Target

Keterangan: Kolom ini merupakan variabel target yang akan diprediksi. Data dalam kolom ini menunjukkan besaran gaji yang ingin diprediksi berdasarkan fitur-fitur lain dalam dataset. Karena bertipe numerik, ini memungkinkan berbagai teknik analisis regresi untuk diterapkan.

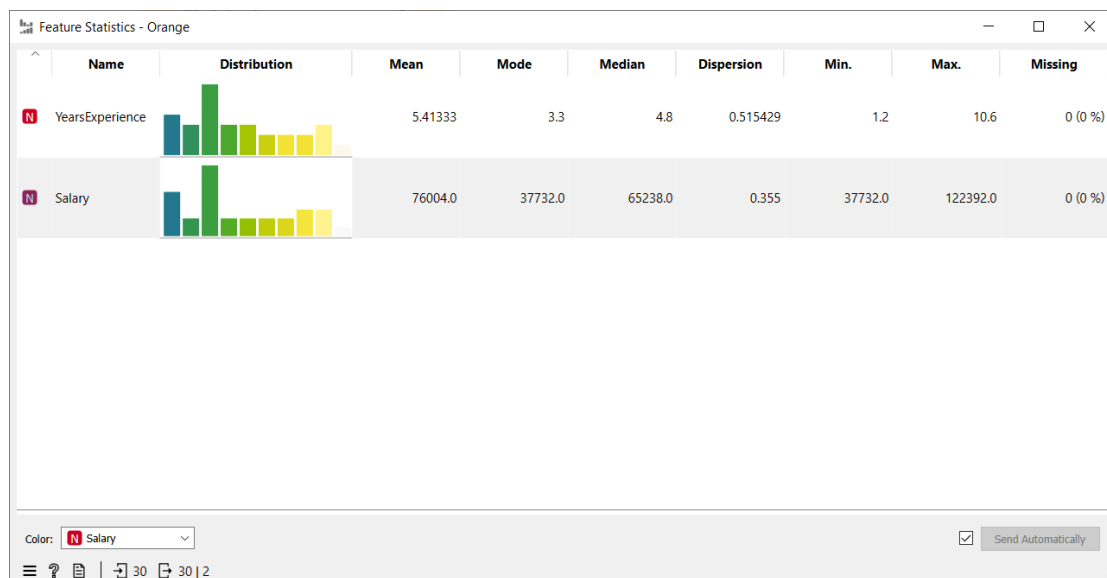
Secara keseluruhan, kolom-kolom ini menyusun struktur dataset yang digunakan untuk analisis gaji, dengan fokus pada prediksi nilai gaji berdasarkan data yang tersedia.



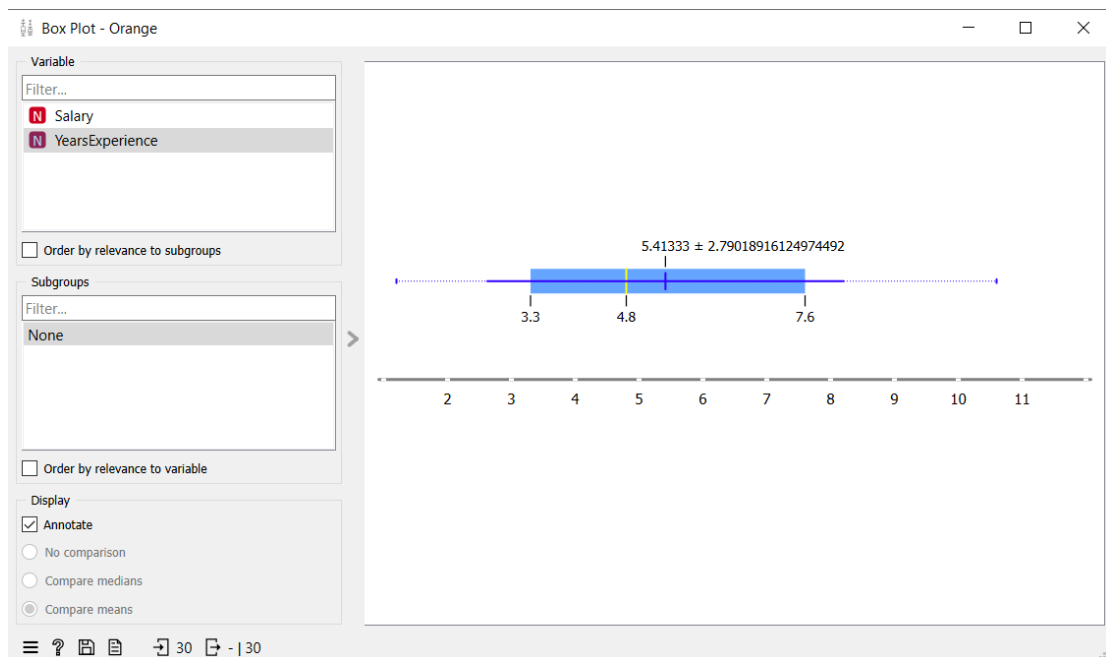
	Salary	YearsExperience
1	39344.0	1.2
2	46206.0	1.4
3	37732.0	1.6
4	43526.0	2.1
5	39892.0	2.3
6	56643.0	3
7	60151.0	3.1
8	54446.0	3.3
9	64446.0	3.3
10	57190.0	3.8
11	63219.0	4
12	55795.0	4.1
13	56958.0	4.1
14	57082.0	4.2
15	61112.0	4.6
16	67939.0	5
17	66030.0	5.2
18	83089.0	5.4
19	81364.0	6
20	93941.0	6.1
21	91739.0	6.9
22	98274.0	7.2

Tampilan tabel data ini berasal dari perangkat lunak analisis data Orange, yang menampilkan 30 instance tanpa data yang hilang. Tabel ini memiliki satu fitur numerik,

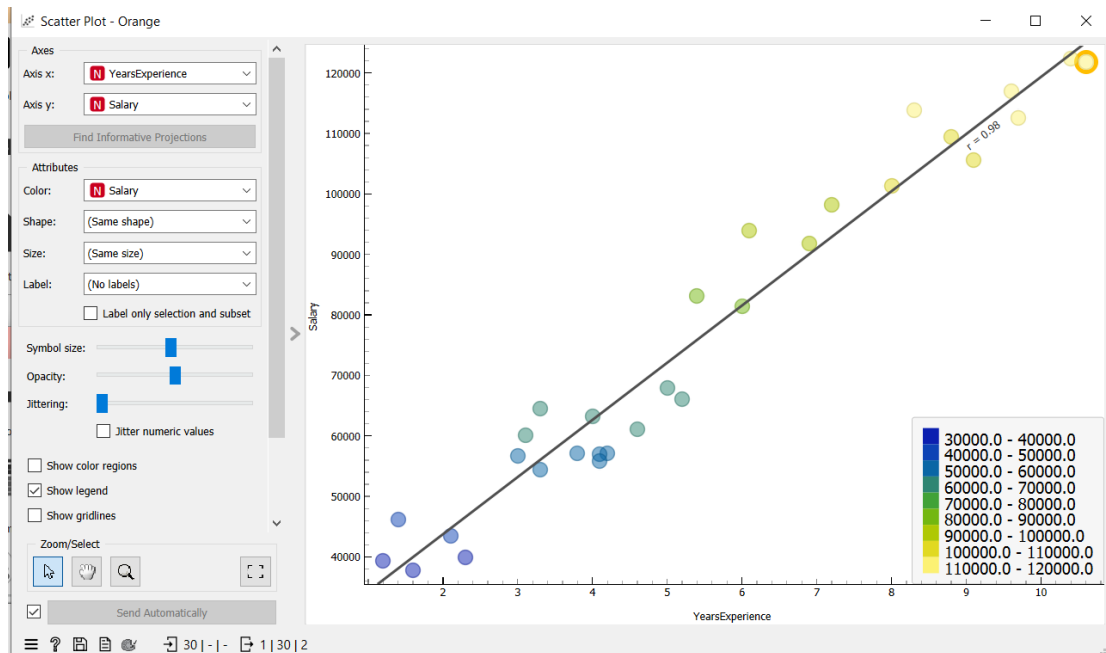
yaitu 'Salary' dan atribut 'YearsExperience', yang menunjukkan hubungan antara pengalaman kerja dan gaji. Setiap baris mewakili satu contoh data, dengan nilai gaji yang tercantum di kolom pertama dan tahun pengalaman di kolom kedua. Pengguna dapat memilih untuk menampilkan label variabel, menyortir nilai numerik, atau mewarnai nilai berdasarkan kategori tertentu untuk analisis visual yang lebih baik. Fitur tambahan seperti "Restore Original Order" dan opsi untuk mengirim data secara otomatis juga tersedia, memudahkan pengguna dalam mengelola dan menganalisis data secara efisien. Tabel ini memberikan gambaran jelas tentang bagaimana data terstruktur dan memungkinkan pengguna untuk melakukan eksplorasi awal terhadap dataset yang digunakan.



Tampilan statistik fitur ini menunjukkan analisis distribusi untuk dua variabel, yaitu 'YearsExperience' dan 'Salary', dalam perangkat lunak analisis data Orange. Setiap variabel dilengkapi dengan histogram yang menggambarkan distribusi frekuensi nilai, memberikan wawasan tentang sebaran data. Selain itu, informasi statistik seperti mean, mode, median, dan dispersion disajikan untuk masing-masing variabel, memungkinkan pengguna untuk memahami karakteristik dasar data. Misalnya, nilai minimum dan maksimum untuk masing-masing variabel juga ditampilkan, bersama dengan jumlah data yang hilang. Dengan demikian, tampilan ini memberikan gambaran menyeluruh tentang distribusi dan sifat statistik dari kedua fitur, yang penting untuk analisis lebih lanjut dan pengambilan keputusan dalam konteks analisis data.



Tampilan box plot ini menunjukkan visualisasi distribusi data untuk variabel 'Salary' dan 'YearsExperience' dalam perangkat lunak analisis data Orange. Box plot adalah alat yang efektif untuk menggambarkan sebaran data, di mana kotak menunjukkan rentang interkuartil (IQR) dan garis di dalam kotak mewakili median. Dengan adanya garis horizontal yang melampaui kotak, pengguna dapat dengan mudah mengidentifikasi outlier dan memahami variasi dalam data. Selain itu, tampilan ini memungkinkan pengguna untuk memfilter dan mengatur data berdasarkan subkelompok tertentu, memberikan fleksibilitas dalam analisis. Dengan demikian, box plot ini memberikan cara yang jelas dan ringkas untuk menganalisis dan membandingkan distribusi nilai antara dua variabel, membantu dalam pengambilan keputusan yang berbasis data.



Tampilan scatter plot ini menggambarkan hubungan antara variabel 'YearsExperience' dan 'Salary' menggunakan perangkat lunak analisis data Orange. Dalam grafik ini, setiap titik mewakili individu, dengan sumbu horizontal menunjukkan jumlah tahun pengalaman kerja dan sumbu vertikal menunjukkan nilai gaji. Garis regresi yang ditambahkan memberikan indikasi tren, menunjukkan bahwa terdapat korelasi positif antara pengalaman kerja dan gaji; semakin banyak tahun pengalaman, semakin tinggi gaji yang diterima. Warna dan ukuran titik juga dapat digunakan untuk menambah dimensi analisis, misalnya, menunjukkan kelompok data tertentu atau memperlihatkan rentang gaji. Dengan demikian, scatter plot ini merupakan alat yang berguna untuk mengidentifikasi pola dan hubungan dalam data, serta untuk membuat prediksi berdasarkan pengalaman kerja dan gaji.